

NARZĘDZIA CLARIN W PRAKTYCE BADAŃ MEDIOZNAWCZYCH

Maria Nowina Konopka

 orcid.org/0000-0001-9978-7621

Instytut Dziennikarstwa, Mediów i Komunikacji Społecznej

Uniwersytet Jagielloński

ABSTRACT

CLARIN Tools in Practice in Media Studies

The CLARIN-PL tools were created for scientists who undertake research in the field of human and social sciences. However, in the field of Polish media studies, the infrastructure of Polish Common Language Resources and Technology is not used, as has been proven by the research based on the analysis of articles published in two Polish media studies journals: *Studia Medioznawcze* and *Zeszyty Prasoznawcze*. The results of the research are presented and discussed in this article which also describes ten tools, applications and services offered by the scientific consortium that may significantly affect both the improvement of the methodology of media research and the comfort of work of Polish scientists.

Keywords: media research, CLARIN, Polish Common Language Resources and Technology Infrastructure, methodology, tools, applications

Wstęp

Analiza popularności języka angielskiego wśród mieszkańców globu prowadzi do interesującej konstatacji. Otóż zgodnie z klasyfikacją portalu *Ethnologue. Languages of the World* język ten jest trzecim najczęściej używanym językiem świata i włada nim na całym globie 1 132 367 000 osób¹. Jego znaczenie wzrasta jednak niewspółmiernie, gdy pole analizy zawężamy prawie dwukrotnie, ograniczając się

1 Portal Ethnologue. Languages of the World [<https://www.ethnologue.com/language/eng>; 29.05.2019].

w badaniach do użytkowników globalnej sieci. W tym przypadku, powołując się na dane *We are social*, należy zaznaczyć, iż aż 54% wszystkich stron internetowych zaadresowanych jest do odbiorców anglojęzycznych (*We are social* 2019)². Warto dodać, iż kolejne miejsca zajmowane są przez język rosyjski i niemiecki – w każdym z nich stworzono jedynie około 6% światowego kontentu. Dysproporcja ta jednoznacznie przesądza o dominującej pozycji *lingua franca* internetu³. Co jednak najważniejsze, w tym języku powstaje nie tylko istotny odsetek światowych zasobów treściowych, ale również oprogramowania i narzędzia do ich analizy. Fakt ten dodatkowo uprzywilejowuje wspomniany język, skutkując implementacją rozwiązań z zakresu najnowszych technologii w języku angielskim także w nieanglojęzycznych krajach europejskich (Nowina Konopka 2018, s. 167–169). W efekcie braku stosownych nakładów finansowych ze strony państwa na wspieranie technologii do przetwarzania języka ojczystego, jak i niskiego poziomu komercyjnej opłacalności tworzenia tego rodzaju narzędzi przez współczesnych gigantów IT (takich jak Apple, Facebook, Google, IBM czy Microsoft), wybrane języki naturalne są stopniowo wypierane z przestrzeni wirtualnej przez powszechnie dostępne, bezpłatne, gotowe, anglojęzyczne rozwiązania informatyczne (Miłkowski 2018). W odpowiedzi na zaistniałą sytuację na początku nowego tysiąclecia zrodziła się ponadpaństwowa inicjatywa CLARIN ERIC, mająca dążyć do niwelowania tych nierówności poprzez wspieranie zaawansowanych technologicznie badań nad językiem naturalnym dla poszczególnych języków europejskich.

Trwające od kilku lat badania i działania wdrożeniowe skutkują stworzeniem kilkuset narzędzi oraz aplikacji do przetwarzania różnych języków, w tym również polskiego (Piasecki 2014). Dzięki nim możliwa staje się praca na wielkich zbiorach danych (*big data*), automatyczne ich analizowanie i przetwarzanie. W Polsce dostępna infrastruktura przeznaczona jest głównie dla naukowców z zakresu nauk społecznych i humanistycznych, a więc także dla środowiska medioznawców. Zastanawiające jest zatem to, na ile w praktyce badań w tym obszarze narzędzia CLARIN są znane i wykorzystywane. Jak się wydaje, proste w zastosowaniu usługi i atrakcyjna wizualizacja wyników analiz nie znajdują zainteresowania wśród badaczy mediów. Celem niniejszego artykułu jest zatem określenie poziomu znajomości wspomnianej infrastruktury badawczej i weryfikacja hipotezy odnoszącej się do przypuszczenia, że narzędzia, aplikacje i usługi CLARIN nie są wykorzystywane przez polskich badaczy w praktyce badań medioznawczych.

2 Według szacunków Davida Nicholasa w roku 2000 odsetek ten sięgał 80% (Nicholas 2001, s. 69).

3 Na dominującą pozycję kultury anglosaskiej wskazuje również Włodzimierz Gogołek, opisując amerykańskie inicjatywy mające na celu poszerzenie zakresu uczestnictwa dóbr kulturowych pochodzących z USA (Gogołek 2010, s. 264–265).

Międzynarodowy projekt CLARIN ERIC

CLARIN ERIC (nazwa jest akronimem powstałym od Common Language Resources and Technology Infrastructure European Research Infrastructure Consortium⁴) to funkcjonująca od 2012 roku ogólnoeuropejska infrastruktura naukowa. Zasadniczym celem projektu jest odpowiedź na zapotrzebowania naukowców coraz częściej podejmujących się prowadzenia badań wymagających pracy na dużych zbiorach danych, tj. artykułach prasowych, wpisach na portalach społecznościowych, tekstach reklamowych, archiwach cyfrowych itp. Nie ma bowiem wątpliwości, że wraz z rozwojem nowych mediów obserwowany jest istotny statystycznie przyrost zasobów treściowych dostępnych w internecie, które współcześnie stały się ważnym i chętnie eksplorowanym materiałem badawczym. W informacyjnym potopie – posługując się terminologią Stanisława Lema – odnalezienie interesującej wiadomości czy tym bardziej poszukiwanie zależności pomiędzy rozproszonymi w sieci zasobami często znacząco przewyższa możliwości poszczególnych badaczy, ukierunkowując ich myślenie na dobór mniejszych prób badawczych (Lem 1996, s. 16–18). W sposób oczywisty zarówno metoda doboru próby, jej wielkość, jak i sposób analizy pozyskanych danych ma zasadniczy wpływ na otrzymane wyniki badań oraz możliwość ich uogólniania. Projekt CLARIN ERIC, korzystający z doświadczeń lingwistyki informatycznej, inżynierii języka naturalnego oraz coraz doskonalszej technologii mowy, zrodził się zatem jako odpowiedź na wyzwania naukowe badaczy XXI wieku, udostępniając im zasoby językowe, elektroniczne narzędzia do samodzielnego i automatycznego przetwarzania języka naturalnego.

Interdyscyplinarny zespół CLARIN zajmuje się budową

narzędzi naukowych, które umożliwią rozszerzony dostęp do dużych zbiorów dokumentów tekstowych, nagrań języka mówionego i zasobów multimedialnych oraz ich zautomatyzowaną analizę. CLARIN pozwala także użytkownikom na przechowywanie źródeł pisanych i mówionych oraz dzielenie się nimi w ramach ogólnoeuropejskiej sieci naukowej. CLARIN dostarcza narzędzi, które umożliwiają wyszukiwanie zasobów i opisywanie ich za pomocą metadanych, a także zapewniają bezpieczny dostęp na prawach licencji określonej przez autora (Clarín_misja).

Europejski projekt zrzesza już przeszło 100 jednostek naukowych rozporozszo-nych na całym kontynencie europejskim, a baczniymi obserwatorami jego rozwoju są również partnerzy z innych części świata, takich jak Stany Zjednoczone czy Republika Południowej Afryki.

4 CLARIC ERIC jest częścią szerszego projektu Europejskiej Mapy Drogowej Infrastruktury Naukowej (ESFRI – European Roadmap for Research Infrastructures, European Strategy Forum on Research Infrastructures).

Polska infrastruktura badawcza technologii językowych dla nauk humanistycznych i społecznych CLARIN – POLSKA

Międzynarodowa działalność CLARIN opiera się na rozproszonej strukturze centrów, które samodzielnie obsługują zasoby językowe i powiązane z nimi usługi oraz aplikacje. Polski węzeł centrum zlokalizowany jest na Politechnice Wrocławskiej, jednakże CLARIN-PL tworzy konsorcjum sześciu jednostek naukowych. Trzonem naukowym są zatem: Grupa Naukowa G4.19 Lingwistyki Informatycznej i Technologii Językowych, Katedra Inteligencji Obliczeniowej Politechniki Wrocławskiej, Zespół Inżynierii Lingwistycznej Instytutu Podstaw Informatyki Polskiej Akademii Nauk, Instytut Slavistyki Polskiej Akademii Nauk, Polsko-Japońska Akademia Technik Komputerowych oraz Instytut Anglistyki Uniwersytetu Łódzkiego (Piasecki 2018).

Omawiana infrastruktura stworzona jest z zasobów, narzędzi i aplikacji udostępnianych wspólnie, ale tworzonych w ramach różnych projektów badawczych. Na przykład „Słowo dnia” (opis bazy w dalszej części tekstu) to projekt realizowany przez pracowników Uniwersytetu Warszawskiego i Fundację Języka Polskiego. Twórcą koncepcji i koordynatorem korpusu „ChronoPress” jest naukowiec związany z Uniwersytetem Wrocławskim, natomiast „Korpusomat” to narzędzie, które integruje funkcjonujące już w ramach grupy CLARIN rozwiązania stworzone przez Zespół Inżynierii Lingwistycznej Instytutu Podstaw Informatyki PAN.

Do podstawowych zadań Centrum Technologii Językowych CL-PL, jak czytamy na oficjalnej stronie internetowej, należy m.in.:

- budowa repozytorium, w którym zgromadzone narzędzia i zasoby zostaną oznaczone trwałymi identyfikatorami;
- dbanie o techniczną spójność powstającego systemu oraz o przestrzeganie przyjętych standardów, praw dotyczących własności intelektualnej, licencji i zasad etycznych;
- ustanowienie polityki bezpieczeństwa, np. poprzez certyfikację serwerów i odpowiedzialne zarządzanie danymi osobowymi (Clarín_projekt).

W praktyce jednak działalność polskiej grupy CLARIN obejmuje, obok tworzenia i udostępniania narzędzi, także dostosowywanie usług i zasobów do indywidualnych potrzeb badaczy będących beneficjentami projektu.

Wykorzystanie narzędzi CLARIN przez autorów polskich czasopism medioznawczych

Celem zweryfikowania postawionej na wstępie hipotezy przeprowadzono badanie przyczynkowe mające na celu określenie poziomu wykorzystania narzędzi CLARIN przez autorów publikujących w polskich czasopismach medioznawczych. Na początkowym etapie do badania wybrano trzy najbardziej rozpoznawalne i najwyżej punktowane czasopisma naukowe skupiające na swych łamach przedstawicieli tego właśnie środowiska. Po analizie procentowego udziału artykułów autorstwa polskich naukowców w stosunku do wszystkich opublikowanych w danym

numerze tekstów z próby postanowiono usunąć periodyk *Central European Journal of Communication*. Wprawdzie grupa CLARIN funkcjonuje na obszarze niemal całej Europy i rozpoznawalność polskiego centrum mogłaby wzrosnąć poprzez jego prezentację dokonaną przez autora spoza Polski, uznano jednak, iż wystarczająca próba tekstów badaczy zagranicznych opublikowana została w dwóch pozostałych czasopismach, czyli *Zeszytach Prasoznawczych* (ZP) i *Studiach Medioznawczych* (SM). Tym samym do analizy skierowano jedynie te czasopisma, w których przeważają rodzimi autorzy, niezależnie od języka, w którym publikują.

Analizę prowadzono od 2015 roku, czyli od momentu uruchomienia polskiego serwisu i udostępnienia go badaczom, do końca 2018 roku. Pod uwagę brano numery czasopism, nie zaś dokładne daty ich wydania. Łącznie we wskazanym okresie w krakowskim czasopiśmie opublikowano 194 artykuły naukowe, a w warszawskim kwartalniku 135. Co warto zaznaczyć, do analizy wybrano jedynie artykuły naukowe, pomijając pozostałe teksty publikowane w takich działach, jak „Recenzje, omówienia, noty” (ZP), „Kronika naukowa” (ZP), „Sprawozdania z konferencji” (SM) i „Recenzje książek” (SM).

W poszukiwaniu badań prowadzonych w oparciu o infrastrukturę CLARIN przeszukano wszystkie 329 artykułów celem zidentyfikowania liczby powołań na źródło danych lub narzędzie analizy. Zasady cytowania zastrzeżone zostały w umowie licencyjnej konsorcjum naukowego CLARIN, a zatem istnieje pewność, iż w przypadku należytego użycia narzędzia czynność ta zostanie odnotowania co najmniej w przypisie (w wybranych przypadkach również w bibliografii).

Wyniki badań potwierdzają przypuszczenie badawcze odnoszące się do niskiej rozpoznawalności narzędzi, jednakże zdumiewające okazuje się pozyskanie wyniku na poziomie zera. Innymi słowy, w żadnej jednostce analizy nie wystąpiło odwołanie do CLARIN czy do dostępnych w jego zasobach narzędzi, usług i aplikacji. Na tej podstawie można wnioskować, że w praktyce badań medioznawczych wspomniane narzędzia nie są wykorzystywane. Interesujące wydaje się pytanie o przyczyny zaistniałego stanu rzeczy, jednakże odpowiedź na nie wymaga badań wykraczających poza przyjęte w metodologii niniejszego tekstu ramy.

Narzędzia, zasoby, usługi i aplikacje CLARIN

Jak zostało już wspomniane, narzędzia znajdujące się w zasobach analizowanej infrastruktury badawczej przeznaczone są głównie dla naukowców z zakresu nauk społecznych, a wiele z nich z powodzeniem wykorzystywać można, prowadząc badania medioznawcze. Z tego też względu w dalszej części tekstu dokonano prezentacji niektórych z nich, mając nadzieję, że przybliżenie ich funkcjonalności przyczyni się do rozpowszechnienia ich wykorzystania. Automatyzacja procesów przetwarzania danych znacząco skraca bowiem procedurę badawczą, eliminując z niej błąd związany z tzw. czynnikiem ludzkim, mogący powstać w wyniku przetwarzania wielkich zasobów danych.

Na internetowej platformie sieci badawczej dostępne usługi podzielone są na dwie główne kategorie: „Zasoby” oraz „Narzędzia i aplikacje”. Pomijając mniej znaczące dla medioznawcy usługi, na szczególną uwagę badaczy w pierwszej grupie zasługują:

- **CHRONOPRESS** – czyli chronologiczny Korpus Polskich Tekstów Prasowych (1945–1954), zawierający ok. 56 tys. fragmentów tekstów prasowych „opracowanych językowo na poziomie morfosyntaktycznym i ustrukturyzowanych pod względem chronologii. Próbkę mają długość ok. 300 wyrazów tekstowych i reprezentują pełne spektrum tematyczne oficjalnego dyskursu publicznego lat 1945–1954 (średnio 12 różnych tytułów gazet lub czasopism na rok)” (Chronopress);
- **SŁOWO DNIA** – to lista słów o wyższej frekwencji w dyskursie medialnym (czyli takich, których frekwencja względna danego dnia jest wyraźnie wyższa niż frekwencja w okresie porównawczym ostatniego roku). Pomiaru dokonuje się codziennie rano na podstawie kanałów RSS wiadomości prasowych siedmiu gazet ogólnopolskich takich jak: *Dziennik Gazeta Prawna*, *Gazeta Wyborcza*, *Polska. The Times*, *Rzeczpospolita*, *Nasz Dziennik*, *Fakt* i *Super Express* (Clarín_słowodnia).

Obraz 1. Zdjęcie strony internetowej projektu „Słowo dnia”

#	Słowo	Komentarz	Kluczowość	Trendy
1	czerwiec	rocznica D-Day, lądowania wojsk alianckich w Normandii (wieczór 6 VI 1944 - wypłynięcie wojsk z Wielkiej Brytanii), inne konteksty	60	↗
2	Macedonia	piątkowy mecz reprezentacji Polski z Macedonią Północną	39	↗
3	transfer	sezon transferów w piłce nożnej i siatkowej, transfery socjalne	22	↗
4	ksiądz	afery, głównie pedofilskie, z udziałem księży	13	↗
5	kamienica	wybuch gazu w kamienicy w Kędzierzynie-Koźlu	11	↗
6	wakacje	zbliżające się wakacje, kierunki wakacyjne, nowy rozkład jazdy PKP	10	↗

Źródło: Portal Słowo dnia [http://słowadnia.clarin-pl.eu/#/default/1927; 29.05.2019]

- **SŁOWOSIEĆ** – to relacyjny słownik semantyczny, odzwierciedlający system leksykalny języka polskiego. W 2019 roku zawierał on 191 tys. rzeczowników, czasowników, przymiotników i przysłówków, 285 tys. znaczeń oraz ponad 600 tys. relacji, będąc tym samym największym wordnetem na świecie (Clarín_wordnet). Pojedyncze znaczenia w polskim wordnecie wzajemnie połączone są relacjami, dzięki czemu powstaje sieć, w której każdy pojedynczy wyraz zdefiniowany jest przez odniesienie do innych wyrazów.

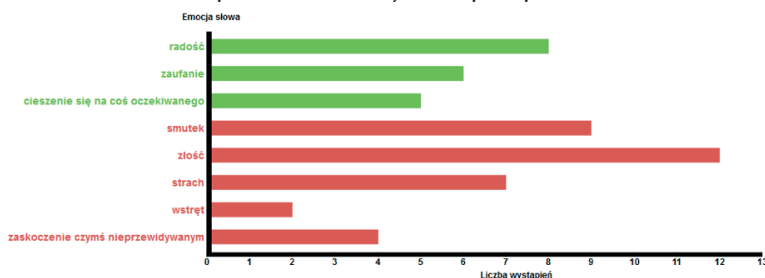
Grupa „Narzędzia i aplikacje” zawiera natomiast takie usługi jak:

- **MOWA** – to interesująca propozycja dla badaczy przeprowadzających wywiady i nagrywających swych rozmówców, a następnie borykających się z problemami ich odsłuchiwania i dokonywania często ręcznej transkrypcji rozmów. Usługa ta pozwala na zautomatyzowanie procesu rozpoznawania mowy i transkrypcji nagrań audio, detekcji mowy i słów kluczowych oraz diaryzacji mówców (podziału sygnału na segmenty mówione przez poszczególnych interlokutorów). W dziale tym bezpłatnie udostępniono ponadto profesjonalne narzędzia wspomagające analizę pozyskanych tekstów (Clarín_mowa);
- **WEBSTY** – to interfejs webowy służący do grupowania tekstów w języku polskim oraz do analizy podobieństwa kilku różnych tekstów (np. celem potwierdzenia autorstwa);
- **TAGER** – służy do tokenizacji tekstu i tagowania morfosyntaktycznego. Dzięki niemu tekst podzielony zostaje na poszczególne słowa, które następnie za pomocą hiperłączy pozwalają przenieść się do innych zasobów, takich jak: synonimy, słowa podobne, tłumaczenia, kolokacje i inne;
- **SUMMARIZE** – jest narzędziem przydatnym każdemu, kto dostrzega u siebie problemy z przygotowywaniem streszczeń własnych artykułów czy innych tekstów (Clarín_summarize). Narzędzie to umożliwia wprowadzenie do programu własnego korpusu i automatyczne sporządzenie wersji skróconej. Opcję tę szczególnie warto polecić osobom dopiero uczącym się przygotowywać artykuły oraz wszystkim tym, którzy chcieliby sprawdzić, jakiego rodzaju podsumowanie przygotowuje za niego program. Należy przy tym zwrócić uwagę na fakt, iż program nie rozumie intencji autora, a jedynie zlicza i przypisuje sens stosowanym przez niego pojęciom. Program może zatem służyć jako forma autoweryfikacji intencji i punkt odniesienia do kontroli faktycznego odwzorowania przyjętych przez autora celów badawczych;
- **SŁOWA KLUCZOWE** – ReSpa – jest narzędziem (podobnie jak to omówione wyżej) pomocnym w pracy z tekstem (Clarín_repsa). Jest ono przeznaczone dla badaczy, których celem jest wyznaczenie słów kluczowych w przygotowanym tekście. Także i w tym przypadku wprowadzenie własnego korpusu oraz wygenerowanie *key words* jest zadaniem technicznie prostym, intuicyjnym i zwracającym wyniki w czasie rzeczywistym.

Ponadto z poziomu innych narzędzi oraz po zalogowaniu się dostępne są:

- **KORPUS DYSKURSU PARLAMENTARNEGO** – będący zbiorem anotowanych lingwistycznie tekstów z posiedzeń Sejmu i Senatu Rzeczypospolitej Polskiej, interpelacji i zapytań poselskich oraz tekstów z posiedzeń komisji, poczynawszy od roku 1919 aż do chwili obecnej (Ogrodniczuk 2018, s. 15–19);
- **WYDŹWIĘK** – czyli narzędzie do zliczania wystąpień w tekście słów o zabarwieniu emocjonalnym. Dzięki łatwemu interfejsowi automatyczna analiza sentymentu daje autorowi podgląd przeprowadzonej analizy celem dokonania ewentualnej ręcznej korekty (Clarín_sentyment).

Wykres 1. Wizualizacja analizy sentymentu



Źródło: dane przykładowe programu wydzźwięk [<http://ws.clarin-pl.eu/sentiment.shtml>; 29.05.2019]

Dziesięć omówionych w niniejszym artykule narzędzi to tylko niektóre z wielu różnorodnych usług dostępnych w ramach CLARIN. Wartością dodaną serwisu jest jego system katalogowy pozwalający na przeglądnięcie z poziomu jednej witryny wszystkich udostępnianych badaczom usług i narzędzi, jak również zapoznanie się z możliwościami stworzonymi dla innych europejskich języków, rozwijanymi przez pozostałe europejskie centra CLARIN ERIC.

Co ważne, infrastruktura przygotowana jest w sposób interoperacyjny, dzięki czemu poszczególne usługi komunikują się ze sobą. Dodatkowo do narzędzi dołączone są zarówno instrukcje i szczegółowe opisy, jak też wersje demonstracyjne oraz repozytorium artykułów i wystąpień konferencyjnych osób tworzących poszczególne narzędzia i korzystających z nich. Grupa CLARIN organizuje również cykliczne szkolenia z zakresu obsługi wybranych narzędzi oraz dostosowuje posiadane usługi do potrzeb konkretnych jednostek badawczych.

Podsumowanie

Międzynarodowy projekt CLARIN ERIC, mający swe centrum także na terytorium Polski, udostępniający zaawansowane technologicznie narzędzia do samodzielnego i automatycznego przetwarzania języka naturalnego (oraz wiele zróżnicowanych zasobów językowych), nie jest w środowisku badaczy z zakresu nauk społecznych powszechnie znany. Analiza artykułów zamieszczanych na łamach dwóch czasopism medioznawczych, tj. *Zeszytów Prasoznawczych* i *Studiów Medioznawczych*, pozwoliła na stwierdzenie, iż polscy badacze w swych analizach w ogóle nie wykorzystują zasobów, usług, aplikacji i narzędzi stworzonych przez polskie konsorcjum naukowo-badawcze. Przekonanie to nie odnosi się jednakże do ogółu prowadzonych przez medioznawców badań, lecz jedynie do tych, których efekty zostały opublikowane we wskazanych periodykach. Wydaje się jednak, że z pewną dozą ostrożności, wnioskując na podstawie przeszło 300 jednostek analizy, konkluzję tę można próbować uogólnić na całość czasopiśmiennictwa medioznawczego. Warto zatem upowszechniać wiedzę na temat dostępnych na stronach CLARIN-PL usług, które, wedle opinii autorki, znacząco wpłynąć mogą zarówno na poprawę warsztatu metodologicznego, jak i na komfort pracy polskich naukowców.

Bibliografia

- Clarín_chronopres. [<http://chronopress.clarin-pl.eu/#!about-corups>; 28.05.2019].
- Clarín_misja. [<http://clarin-pl.eu/pl/misja>; 28.05.2019].
- Carin_mowa. [<http://mowa.clarin-pl.eu/tools>; 29.05.2019].
- Clarín_respa. [<http://ws.clarin-pl.eu/respa.shtml>; 28.05.2019].
- Clarín_sentyment. [<http://ws.clarin-pl.eu/sentiment.shtml>; 28.05.2019].
- Clarín_słownia. [<http://slowadnia.clarin-pl.eu/#/default>; 29.05.2019].
- Clarín_summarize. [<http://ws.clarin-pl.eu/summarize.shtml>; 28.05.2019].
- Clarín_wordnet. [<http://plwordnet.pwr.wroc.pl/wordnet/about>; 28.05.2019].
- Ethnologue. Languages of the World. [<https://www.ethnologue.com>; 29.05.2019].
- Gogolek W. (2010). Komunikacja sieciowa. Uwarunkowania, kategorie i paradoksy. Warszawa.
- Lem S. (1996). Ryzyko Internetu, *PC Magazine*, nr 5, s. 16–18.
- Miłkowski M. (2018). Język polski w erze cyfrowej. [<http://www.meta-net.eu/whitepapers/e-book/polish.pdf>; 1.05.2019].
- Nicholas D. (2001). Ocena potrzeb informacyjnych w dobie Internetu. Idee, metody, środki, tłum. J. Woźniak, M. Kisilowska. Warszawa.
- Nowina Konopka M. (2018). Infomorfoza. Zarządzanie informacją w nowych mediach. Kraków.
- Ogrodniczuk M. (2018). Polish Parliamentary Corpus. W: D. Fišer, M. Eskevich, F. de Jong (ed.). Proceedings of the LREC 2018 Workshop ParlaCLARIN: Creating and Using Parliamentary Corpora (p. 15–19). Paris.
- Piasecki M. (2014). User-driven Language Technology Infrastructure – the Case of CLARIN-PL, 9th Language Technologies Conference Information Society – IS 2014. [http://nl.ijs.si/isjt14/proceedings/isjt2014_01.pdf; 10.05.2019].
- Piasecki M. (2018). CLARIN-PL – wielka infrastruktura badawcza technologii językowych dla nauk humanistycznych i społecznych. [<http://clarin-pl.eu/wp-content/uploads/2018/07/CLARIN-PL-wprowadzenie.pdf>; 19.05.2019].
- We are social, Global Digital Report. 2018. [<https://wearesocial.com/global-digital-report-2019>; 7.05.2019].

STRESZCZENIE

Narzędzia CLARIN-PL stworzone zostały z myślą o naukowcach podejmujących się badań z obszaru nauk społecznych i humanistycznych. W praktyce badań medioznawczych polska infrastruktura badawcza technologii językowych nie jest jednak wykorzystywana, co udowodniono, przeprowadzając badania w oparciu o analizę artykułów publikowanych w dwóch polskich czasopismach medioznawczych: *Studiach Medioznawczych* i *Zeszytach Prasoznawczych*. W tekście opisano również dziesięć oferowanych przez konsorcjum naukowe narzędzi, aplikacji i usług, mogących, w ocenie autorki, znacząco wpłynąć zarówno na poprawę warsztatu metodologicznego, jak i na komfort pracy polskich naukowców.

Słowa kluczowe: badania medioznawcze, CLARIN, polska infrastruktura badawcza technologii językowych, metodologia, narzędzia, aplikacje